Propagation of Memory Parameter from Durations to Counts

Rohit Deo* Clifford M. Hurvich* Philippe Soulier† Yi Wang*

February 2, 2008

Abstract

We establish sufficient conditions on durations that are stationary with finite variance and memory parameter $d \in [0, 1/2)$ to ensure that the corresponding counting process $N(t)$ satisfies $\operatorname{Var} N(t) \sim Ct^{2d+1}$ $(C > 0)$ as $t \to \infty$, with the same memory parameter $d \in [0, 1/2)$ that was assumed for the durations. Thus, these conditions ensure that the memory in durations propagates to the same memory parameter in counts and therefore in realized volatility. We then show that any Autoregressive Conditional Duration ACD(1,1) model with a sufficient number of finite moments yields short memory in counts, while any Long Memory Stochastic Duration model with $d > 0$ and all finite moments yields long memory in counts, with the same d. Finally, we present a result implying that the only way for a series of counts aggregated over a long time period to have nontrivial autocorrelation is for the short-term counts to have long memory. In other words, aggregation ultimately destroys all autocorrelation in counts, if and only if the counts have short memory.

KEYWORDS: Long Memory Stochastic Duration, Autoregressive Conditional Duration, Rosenthal-type Inequality.

*New York University, 44 W. 4'th Street, New York NY 10012 USA

†Université Paris X, 200 avenue de la République, 92001 Nanterre cedex, France

The authors thank Xiaohong Chen and Raymond Thomas for helpful comments and suggestions.

I Introduction

There is a growing literature on long memory in volatility of financial time series. See, e.g., Robinson (1991), Bollerslev and Mikkelsen (1996), Robinson and Henry (1999), Deo and Hurvich (2001), Hurvich, Moulines and Soulier (2005). Long memory in volatility, which has been repeatedly found in the empirical literature, plays a key role in the forecasting of realized volatility (Andersen, Bollerslev, Diebold and Labys 2001, Deo, Hurvich and Lu 2005), and has important implications on option pricing (see Comte and Renault 1998).

Given the increasing availability of transaction-level data it is of interest to explain phenomena observed at longer time scales from equally-spaced returns in terms of more fundamental properties at the transaction level. Engle and Russell (1998) proposed the Autoregressive Conditional Duration (ACD) model to describe the durations between trades, and briefly explored the implications of this model on volatility of returns in discrete time, though they did not determine the persistence of this volatility, as measured, say, by the decay rate of the autocorrelations of the squared returns. Deo, Hsieh and Hurvich (2005) proposed the Long-Memory Stochastic Duration (LMSD) model, and began an empirical and theoretical exploration of the question as to which properties of durations lead to long memory in volatility, though the theoretical results presented there were not definitive.

The collection of time points $\cdots t_{-1} < t_0 \leq 0 < t_1 < t_2 < \cdots$ at which a transaction (say, a trade of a particular stock on a specific market) takes place, comprises a point process, a fact which was exploited by Engle and Russell (1988). These event times $\{t_k\}$ determine a *counting process*,

$$N(t) = Number\ of\ Events\ in\ (0, t].$$

For any fixed time spacing $\Delta t > 0$, one can define the *counts* $\Delta N_{t'} = N(t'\Delta t) - N((t'-1)\Delta t)$, the number of events in the t''th time interval of width Δt, where $t' = 1, 2, \cdots$. The event times $\{t_k\}_{k=-\infty}^{\infty}$ also determine the *durations*, given by $\{\tau_k\}_{k=-\infty}^{\infty}$, $\tau_k = t_k - t_{k-1}$.

Both the ACD and LMSD models imply that the doubly infinite sequence of durations $\{\tau_k\}_{k=-\infty}^{\infty}$ are a stationary time series, i.e., there exists a probability measure P^0 under which the joint distribution of

any subcollection of the $\{\tau_k\}$ depends only on the lags between the entries. On the other hand, a point process N on the real line is stationary under the measure P if $P(N(A)) = P(N(A+c))$ for all real c. A fundamental fact about point processes is that in general (a notable exception is the Poisson process) there is no single measure under which both the point process N and the durations $\{\tau_k\}$ are stationary, i.e., in general P and P^0 are not the same. Nevertheless, there is a one-to-one correspondence between the class of measures P^0 that determine a stationary duration sequence and the class of measures P that determine a stationary point process. The measure P^0 corresponding to P is called the *Palm distribution*. The counts are stationary under P, while the durations are stationary under P^0.

Deo, Hsieh and Hurvich (2005) pointed out, using a theorem of Daley, Rolski and Vesilo (2000) that if durations are generated by an ACD model and if the durations have tail index $\kappa \in (1,2)$ under P^0, then the resulting counting process $N(t)$ has long range count dependence with memory parameter $d \geq 1-\kappa/2$, in the sense that $\mathrm{Var}\, N(t) \sim Cn^{1+2d}$ $(C > 0)$ as $t \to \infty$, under P. This, together with the model for returns at equally spaced intervals of time given in Deo, Hsieh and Hurvich (2005) implies that realized volatility has long memory in the sense that the n-term partial sum of realized volatility has a variance that scales as $C_2 n^{2d+1}$ as $n \to \infty$, where $C_2 > 0$. Deo, Hsieh and Hurvich (2005) also showed that if durations are generated by an LMSD model with memory parameter d under P^0 then counts have long memory with memory parameter $d^{counts} \geq d$, but unfortunately this conclusion was established only under the duration-stationary measure P^0, and not under the count-stationary measure P. This gap can be bridged using methods described in this paper. Still, the results we have described above merely give lower bounds for the memory parameter in counts.

In this paper, we will establish sufficient conditions on durations that are stationary with finite variance and memory parameter $d \in [0, 1/2)$ under P^0 to ensure that the corresponding counting process $N(t)$ satisfies $\mathrm{Var}\, N(t) \sim Ct^{2d+1}$ $(C > 0)$ as $t \to \infty$ under P, with the same memory parameter $d \in [0, 1/2)$ that was assumed for the durations. Thus, these conditions ensure that the memory in durations propagates to the same memory parameter in counts and therefore in realized volatility.

Next, we will verify that the sufficient conditions of our Theorem 1 are satisfied for the ACD(1,1)

model assuming finite $8+\delta$ moment ($\delta > 0$) of the durations under P^0, and for the LMSD model with any $d \in [0, 1/2)$ assuming that the multiplying shocks have all moments finite. Thus, any ACD(1,1) model with a sufficient number of finite moments yields short memory in counts, while any LMSD model with $d > 0$ and all finite moments yields long memory in counts. These results for the LMSD and ACD(1,1) models are given in Theorems 2 and 3, respectively. Lemma 1, which is used in proving Theorem 2, provides a Rosenthal-type inequality for moments of absolute standardized partial sums of durations under the LMSD model, and is of interest in its own right.

Finally, we present a result (Theorem 4) implying that if counts have memory parameter $d \in [0, 1/2)$ then further aggregations of these counts to longer time intervals will have a lag-1 autocorrelation that tends to $2^{2d} - 1$ as the level of aggregation grows. Interestingly, this limit is zero if and only if $d = 0$. Thus, one of the important functions of long memory in counts is that it allows the counts to have a non-vanishing autocorrelation even as Δt grows, as was found by Deo, Hsieh and Hurvich (2005) to occur in empirical data. By contrast, short memory in counts implies that counts at long time scales (large Δt) are essentially uncorrelated, in contradiction to what is seen in actual data. To summarize, aggregation ultimately destroys all autocorrelation in counts, if and only if the counts have short memory.

II Theorems on the propagation of the memory parameter

Let E, E^0, Var, Var^0 denote expectations and variances under P and P^0, respectively. Define $\mu = E^0(\tau_k)$ and $\lambda = \frac{1}{\mu}$. Our main theorem uses the assumption that P^0 is $\{\tau_k\}$-mixing, defined as follows. Let $\mathcal{N} = \sigma(\{\tau_k\}_{k=-\infty}^{\infty})$ and $\mathcal{F}_n = \sigma(\{\tau_k\}_{k=n}^{\infty})$. We say that P^0 is $\{\tau_k\}$-mixing if

$$\lim_{n\to\infty} \sup_{B \in \mathcal{N}\cap\mathcal{F}_n} |P^0(A \cap B) - P^0(A)P^0(B)| = 0$$

for all $A \in \mathcal{N}$.

Theorem 1 *Let $\{\tau_k\}$ be a duration process such that the following conditions hold:*

i) $\{\tau_k\}$ is stationary under P^0.

ii) P^0 *is* $\{\tau_k\}$*-mixing.*

iii) $\exists\ d \in [0, \frac{1}{2})$ *such that*

$$Y_n(s) = \frac{\sum_{k=1}^{\lfloor ns \rfloor}(\tau_k - \mu)}{n^{1/2+d}}, \qquad s \in [0,1]$$

converges weakly to $\sigma B_{1/2+d}(\cdot)$ *under* P^0*, where* $\sigma > 0$ *and* $B_{1/2+d}(\cdot)$ *is fractional Brownian motion if* $0 < d < \frac{1}{2}$ *or standard Brownian motion* $B_{1/2} = B$ *if* $d = 0$*.*

iv)

$$\sup_n E^0 \Big| \frac{\sum_{k=1}^{n}(\tau_k - \mu)}{n^{1/2+d}} \Big|^p < \infty \qquad \begin{cases} for\ all\ p > 0,\ if\ d \in (0, \frac{1}{2}) \\ for\ p = 8 + \delta,\ \delta > 0,\ if\ d = 0 \end{cases} .$$

Then the induced counting process $N(t)$ *satisfies* $VarN(t) \sim Ct^{2d+1}$ *under* P *as* $t \to \infty$ *where* $C > 0$*.*

Remark: Inspection of the proof of Theorem 1 reveals that if $d > 0$, only $4/(0.5 - d) + \delta$ finite moments are needed, where $\delta > 0$ is arbitrarily small. The closer d is to 1/2, the larger the number of finite moments required.

Remark: As pointed out by Nieuwenhuis (1989), if $\{\tau_k\}$ is strong mixing under P^0 then P^0 is $\{\tau_k\}$-mixing. This weaker form of mixing is essential for our purposes since even Gaussian long-memory processes are not strong mixing. See Guégan and Ladoucette (2001).

A LMSD Process

Define the LMSD process $\{\tau_k\}_{k=-\infty}^{\infty}$ for $d \in [0, \frac{1}{2})$ as

$$\tau_k = e^{h_k}\epsilon_k$$

where under P^0 the $\epsilon_k \geq 0$ are *i.i.d.* with all moments finite, and $h_k = \sum_{j=0}^{\infty} b_j e_{k-j}$, the $\{e_k\}$ are *i.i.d.* Gaussian with zero mean, independent of $\{\epsilon_k\}$, and

$$b_j \sim \begin{cases} Cj^{d-1} & if\ d \in (0, \frac{1}{2}) \\ Ca^j,\ |a| < 1 & if\ d = 0 \end{cases}$$

($C \neq 0$) as $j \to \infty$. Note that for convenience, we nest the short-memory case ($d = 0$) within the LMSD model, so that the allowable values for d in this model are $0 \leq d < 1/2$.

Theorem 2 *If the durations $\{\tau_k\}$ are generated by the LMSD process with $d \in [0, 1/2)$, then the induced counting process $N(t)$ satisfies $VarN(t) \sim Ct^{2d+1}$ under P as $t \to \infty$ where $C > 0$.*

To establish Theorem 2, we will use the following Rosenthal-type inequality.

Lemma 1 *For durations $\{\tau_k\}$ generated by the LMSD process with $d \in [0, \frac{1}{2})$, for any fixed positive integer $p \geq 2$, $E^0\{|y_n - E^0(y_n)|^p\}$ is bounded uniformly in n, where*

$$y_n = \frac{\sum_{k=1}^{n} \tau_k}{n^{1/2+d}} \qquad .$$

B ACD(1,1) Process

Define the ACD(1,1) process $\{\tau_k\}_{k=-\infty}^{\infty}$ as

$$\begin{aligned} \tau_k &= \psi_k \epsilon_k \\ \psi_k &= \omega + \alpha \tau_{k-1} + \beta \psi_{k-1} \end{aligned}$$

with $\omega > 0, \alpha > 0, \beta \geq 0$ and $\alpha + \beta < 1$, where under P^0, $\epsilon_k \geq 0$ are *i.i.d.* with mean 1. We will assume further that under P^0, ϵ_k has a density g_ϵ such that $\int_0^\theta g_\epsilon(x)dx > 0, \forall\, \theta > 0$ and $E^0(\tau_k^{8+\delta}) < \infty$ for some $\delta > 0$.

Nelson (1990) guarantees the existence of the doubly-infinite ACD(1,1) process $\{\tau_k\}_{k=-\infty}^{\infty}$, which in our terminology is stationary under P^0.

Theorem 3 *Suppose that the durations $\{\tau_k\}$ are generated by the ACD(1,1) model, with the additional assumptions stated above. Then the induced counting process $N(t)$ satisfies $VarN(t) \sim Ct$ under P as $t \to \infty$ where $C > 0$.*

III Autocorrelation of Aggregated Counts

Theorem 4 *Let $\{X_t\}$ be a stationary process such that $Var(\sum_{t=1}^n X_t) \sim Cn^{1+2d}$ as $n \to \infty$, where $C \neq 0$ and $d \in [0, 1/2)$. Then*

$$\lim_{n\to\infty} Corr\left[\sum_{t=1}^{n} X_t, \sum_{t=n+1}^{2n} X_t\right] = 2^{2d} - 1.$$

Proof:

$$\text{Var}\left[\sum_{t=1}^{2n} X_t\right] = 2\,\text{Var}\left[\sum_{t=1}^{n} X_t\right] + 2\,\text{Cov}\left[\sum_{t=1}^{n} X_t, \sum_{t=n+1}^{2n} X_t\right].$$

Thus,

$$\text{Cov}\left[\sum_{t=1}^{n} X_t, \sum_{t=n+1}^{2n} X_t\right] = .5\left(\text{Var}\left[\sum_{t=1}^{2n} X_t\right] - 2\text{Var}\left[\sum_{t=1}^{n} X_t\right]\right).$$

The result follows by noting that $\lim_{n\to\infty} n^{-2d-1}\text{Var}(\sum_{t=1}^n X_t) = C$, where $C \neq 0$. $\square$

This theorem has an interesting practical interpretation. If we write $X_k = N[k\Delta t] - N[(k-1)\Delta t]$ where $\Delta t > 0$ is fixed, then X_k represents the number of events (count) in a time interval of width Δt, e.g. one minute. Thus, $\sum_{k=1}^n X_k$ is the number of events in a time interval of length n minutes, e.g. one day. The theorem implies that as the level of aggregation (n) increases, the lag-1 autocorrelation of the aggregated counts will approach a nonzero constant if and only if the non-aggregated count series $\{X_k\}$ has long memory. In other words, the only way for a series of counts over a long time period to have nontrivial autocorrelation is for the short-term counts to have long memory. Since in practice long-term counts do have substantial autocorrelation (see Deo, Hsieh and Hurvich 2005), it is important to use only the models for durations that imply long memory in the counting process (LRcD). Examples of such models include the LMSD model (see Theorem 2), and ACD models with infinite variance (see Daley, Rolski and Vesilo (2000), and Theorem 2 of Deo, Hsieh and Hurvich, 2005).

IV Appendix: Proofs

Let P denote the stationary distribution of the point process N on the real line, and let P^0 denote the corresponding Palm distribution. P determines and is completely determined by the stationary distribution P^0 of the doubly infinite sequence $\{\tau_k\}_{k=-\infty}^{\infty}$ of durations. Note that the counting process N is stationary under P, the durations are stationary under P_0, but in general there is no single distribution under which both the counting process and the durations are stationary. For more details on the correspondence between P and P^0, see Daley and Vere-Jones (2003), Baccelli and Brémaud (2003), or Nieuwenhuis (1989).

Following the standard notation for point processes on the real line (see, e.g., Nieuwenhuis 1989, p. 594), we assume that the event times $\{t_k\}_{k=-\infty}^{\infty}$ satisfy

$$\ldots < t_{-1} < t_0 \leq 0 < t_1 < t_2 < \ldots .$$

Let

$$u_k = \begin{cases} t_1 & if\ k = 1 \\ \tau_k & if\ k \geq 2 \end{cases} .$$

Here, the random variable $t_1 > 0$ is the time of occurrence of the first event following $t = 0$. For $t > 0$, define the count on the interval $(0, t]$, $N(t) := N(0, t]$, by

$$\begin{aligned} N(t) &= \max\{s : \sum_{i=1}^{s} u_i \leq t\}, \quad u_1 \leq t \\ &= 0, \quad\quad\quad\quad\quad\quad\quad\quad u_1 > t. \end{aligned}$$

Throughout the paper, the symbol $\Longrightarrow$ denotes weak convergence in the space $D[0, 1]$.

Proof of Theorem 1:

By assumption $iii)$, $Y_n \Longrightarrow \sigma B_{1/2+d}$ under P^0, where $\sigma > 0$. First, we will apply Theorem 6.3 of Nieuwenhuis (1989) to the durations $\{\tau_k\}_{k=-\infty}^{\infty}$ to conclude that $Y_n \Longrightarrow \sigma B_{1/2+d}$ under P. Since the $\{\tau_k\}_{k=-\infty}^{\infty}$ are stationary under P^0 and are generated by the shift to the first event following time zero

(see Nieuwenhuis 1989, p. 600), and since we have assumed that P^0 is $\{\tau_k\}$-mixing, his Theorem 6.3 applies. It follows that $Y_n \Longrightarrow \sigma B_{1/2+d}$ under P. We next show that the suitably normalized counting process converges to the same limit under P.

Define

$$\tilde{Y}_n(s) = \frac{\sum_{k=1}^{\lfloor ns \rfloor}(u_k - \mu)}{n^{1/2+d}} \quad , \quad s \in [0, 1].$$

Note that for all s, $\tilde{Y}_n(s) = Y_n(s) + n^{-(1/2+d)}(u_1 - \tau_1)$. From Baccelli and Brémaud (2003, Equation 1.4.2, page 33), for any measurable function h,

$$E[h(\tau_1)] = \lambda E^0[\tau_1 h(\tau_1)] \quad . \tag{1}$$

Since $u_1 \leq \tau_1$, and since assumption iv) implies that τ_1 has finite variance under P^0, using $h(x) = x$ in (1), it follows that $n^{-(1/2+d)}(u_1 - \tau_1)$ is $o_p(1)$ under P. Thus, $\tilde{Y}_n \Longrightarrow \sigma B_{1/2+d}$ under P.

Let

$$Z(t) = \frac{N(t) - t/\mu}{t^{1/2+d}} \quad . \tag{2}$$

By Iglehart and Whitt (1971, Theorem 1), it follows that $Z(t) \xrightarrow{d} \tilde{C} B_{1/2+d}(1)$ under P as $t \to \infty$, where $\tilde{C} > 0$. Furthermore, by Lemma 2, $Z^2(t)$ is uniformly integrable under P and hence $\lim_t \text{Var}[Z(t)] = \tilde{C}^2 \text{Var}[B_{1/2+d}(1)]$. The theorem is proved. $\square$

Proof of Theorem 2:

We simply verify that the conditions of Theorem 1 hold for this process.

By definition $\{\tau_k\}$ is stationary under P^0 and by Lemma 4, P^0 is $\{\tau_k\}$ mixing. By Surgailis and Viano (2002), $Y_n \Longrightarrow \sigma B_{1/2+d}$ under P^0, where $\sigma > 0$ and by Lemma 1, $\sup_n E^0 \left| \frac{\sum_{k=1}^{n}(\tau_k - \mu)}{n^{1/2+d}} \right|^p < \infty$ for all p. Thus, the result is proved. $\square$

Proof of Theorem 3:

We simply verify that the conditions of Theorem 1 hold for this process.

By Lemma 4, $\{\tau_k\}$ is exponential α-mixing, and hence strong mixing and thus by Nieuwenhuis (1989), P^0 is $\{\tau_k\}$-mixing. Furthermore, since all moments of τ_k exist up to order $8+\delta, \delta > 0$, we can apply results from Doukhan (1994) to obtain

$$Y_n \Rightarrow CB, \tag{3}$$

if $\frac{1}{n}\mathrm{var}(\sum_{k=1}^{n}\tau_k) \to C^2 > 0$, as $n \to \infty$.

It is well known that the GARCH(1,1) model can be represented as an ARMA(1,1) model, see Tsay (2002). Similarly, the ACD(1,1) model can also be re-formulated as an ARMA(1,1) model,

$$\tau_k = \omega + (\alpha+\beta)\tau_{k-1} + (\eta_k - \beta\eta_{k-1}) \tag{4}$$

where $\eta_k = \tau_k - \psi_k$ is white noise with finite variance since $E(\tau_k^{8+\delta}) < \infty$. The autoregressive and moving average parameters of the resulting ARMA(1,1) model are $(\alpha+\beta)$ and β, respectively.

It is also known that for any stationary invertible ARMA model $\{z_k\}$, $n\mathrm{var}(\bar{z}) \to 2\pi f_z(0)$, where $f_z(0)$ is the spectral density of $\{z_k\}$ at zero frequency. For an ARMA(1,1) process, $f_z(0) > 0$ if the moving average coefficient is less than 1. Here, since $0 \le \beta < 1$, we obtain $\frac{1}{n}\mathrm{var}(\sum_{k=1}^{n}\tau_k) = n\mathrm{var}(\bar{\tau}) \to 2\pi f_\tau(0) > 0$, as $n \to \infty$. Therefore (3) follows.

Define $y_n = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\tau_k$. Since all moments of τ_k are bounded up to order $8+\delta$, $(\delta > 0)$ under P^0, by Yokoyama (1980), we obtain

$$E^0\{|y_n - E^0(y_n)|^{8+\delta}\} \le K < \infty, \qquad \delta > 0 \tag{5}$$

uniformly in n, provided that $\{\tau_k\}$ is exponential α-mixing, which is proved in Lemma 4.

Therefore, we can apply Theorem 1 to the ACD(1,1) model and the result follows. $\square$

Proof of Lemma 1:

We present the proof for the case $0 < d < \frac{1}{2}$. The proof for the case $d = 0$ follows along similar lines. Also, we assume here that p is a positive even integer. The result for all positive odd integers follows by Hölder's inequality.

Let $\tilde{y}_n = y_n - E^0(y_n)$. Since $p \geq 2$ is even and $E^0(\tilde{y}_n)^p$ can be expressed as a linear combination of the products of the joint cumulants of $\tilde{y}_n$ of order $2, \ldots, p$, we have

$$\begin{aligned} 0 \leq E^0|\tilde{y}_n|^p = E^0(\tilde{y}_n^p) &= \sum_{\pi} \Big[c_\pi \prod_{j \in \pi} \mathrm{cum}(\underbrace{\tilde{y}_n, \ldots, \tilde{y}_n}_{j \text{ terms}}) \Big] \\ &\leq \sum_{\pi} \Big[|c_\pi| \prod_{j \in \pi} |\mathrm{cum}(\underbrace{\tilde{y}_n, \ldots, \tilde{y}_n}_{j \text{ terms}})| \Big] \end{aligned}$$

where π ranges over the additive partitions of n and c_π is a finite constant depending on π.

Since the first order cumulant of $\tilde{y}_n$ is zero and for all integers $m \geq 2$, the m-th order cumulant of $\tilde{y}_n$ is equal to that of y_n, it suffices to show that the absolute value of the m-th order cumulant of y_n is bounded uniformly in n under P^0, for all $m \in \{2, \ldots, p\}$.

We first consider the second and the third order cumulants.

For the second order cumulant ($m = 2$),

$$|\mathrm{cum}(y_n, y_n)| = |\mathrm{cum}(\frac{\sum_{k=1}^n \tau_k}{n^{d+\frac{1}{2}}}, \frac{\sum_{s=1}^n \tau_s}{n^{d+\frac{1}{2}}})| \leq \frac{1}{n^{2d+1}} \sum_{k=1}^n \sum_{s=1}^n |\mathrm{cum}(\tau_k, \tau_s)| \qquad .$$

To calculate the joint cumulant $\mathrm{cum}(\tau_k, \tau_s)$, we briefly introduce some terminology, mainly cited from Brillinger (1981): consider a (not necessary rectangular) two-way table of indices,

$$\begin{array}{ccc} (1,1) & \ldots & (1, J_1) \\ \vdots & \ldots & \vdots \\ (I,1) & \ldots & (I, J_I) \end{array}$$

and a partition $P_1 \cup P_2 \cup \ldots \cup P_M$ of its entries. We say sets $P_{m'}$, $P_{m''}$ of the partition **hook** if there exist $(i_1, j_1) \in P_{m'}$ and $(i_2, j_2) \in P_{m''}$ such that $i_1 = i_2$, i.e. at least one entry of $P_{m'}$ and one entry of $P_{m''}$ come from the same row in the two-way table. We say that sets $P_{m'}$ and $P_{m''}$ **communicate** if there exists a sequence of sets $P_{m_1} = P_{m'}, P_{m_2}, \ldots, P_{m_N} = P_{m''}$ such that P_{m_n} and $P_{m_{n+1}}$ hook for $n = 1, \ldots, N-1$. So $P_{m'}$ and $P_{m''}$ communicate as long as one can find an ordered sequence of sets such that all neighboring pairs hook, and this sequence links $P_{m'}$ and $P_{m''}$ together. Finally a partition is said to be **indecomposable** if all sets in the partition communicate.

By Brillinger (1981), Theorem 2.3.2, for a two-way array of random variables X_{ij}, $j = 1, \ldots, J_i$, $i = 1, \ldots, I$ (see the corresponding two-way table above), the joint cumulant of the I row products

$$Y_i = \prod_{j=1}^{J_i} X_{ij}, \qquad i = 1, \ldots, I$$

is given by,

$$\mathrm{cum}(Y_1, \ldots, Y_I) = \sum_{\nu} \mathrm{cum}(X_{ij}; ij \in \nu_1) \ldots \mathrm{cum}(X_{ij}; ij \in \nu_w)$$

where the summation is over all indecomposable partition $\nu = \nu_1 \cup \ldots \cup \nu_w$ of the two-way table of indices.

It is more convenient to write the partitions in terms of symbols representing the random variables, instead of the indices themselves. We will always use distinct symbols, so that there is a one-to-one correspondence between the indices and the symbols. Nevertheless, the random variables represented by distinct symbols need not be distinct. For example, e^{h_k} and e^{h_s} are distinct symbols, but if $k = s$, they are not different random variables. Ultimately, the cumulants are computed from the random variables.

To compute $\mathrm{cum}(\tau_k, \tau_s)$, we use the two-way table of indices (left) and the corresponding table of symbols (right),

$$\begin{array}{cc} (1,1) & (1,2) \\ (2,1) & (2,2) \end{array} \quad , \quad \begin{array}{cc} e^{h_k} & \epsilon_k \\ e^{h_s} & \epsilon_s \end{array}$$

with $I = 2, J_1 = 2$ and $J_2 = 2$.

From Brillinger (1981), Theorem 2.3.1, all joint cumulants corresponding to partitions with at least one of the symbols representing $\{e^{h_k}\}$ and at least one of the symbols representing $\{\epsilon_k\}$ in the same set, are zero because the corresponding random variable sequences are mutually independent. So for $m = 2$, excluding those with at least one of e^{h_k}, e^{h_s} and at least one of ϵ_k, ϵ_s in the same set, the only possible indecomposable partitions (here, the partition is given in terms of the symbols) are:

$$\{e^{h_k}, e^{h_s}\}, \{\epsilon_k, \epsilon_s\}$$

$$\{e^{h_k}, e^{h_s}\}, \{\epsilon_k\}, \{\epsilon_s\}$$

$$\{e^{h_k}\}, \{e^{h_s}\}, \{\epsilon_k, \epsilon_s\} \qquad .$$

Thus, $|\mathrm{cum}(y_n, y_n)| \leq A + B + C$, where,

$$\begin{aligned}
A &= \frac{1}{n^{2d+1}} \sum_{k=1}^{n} \sum_{s=1}^{n} |\mathrm{cum}(e^{h_k}, e^{h_s})||\mathrm{cum}(\epsilon_k, \epsilon_s)| \\
B &= \frac{1}{n^{2d+1}} \sum_{k=1}^{n} \sum_{s=1}^{n} |\mathrm{cum}(e^{h_k})\mathrm{cum}(e^{h_s})||\mathrm{cum}(\epsilon_k, \epsilon_s)| \\
C &= \frac{1}{n^{2d+1}} \sum_{k=1}^{n} \sum_{s=1}^{n} |\mathrm{cum}(e^{h_k}, e^{h_s})||\mathrm{cum}(\epsilon_k)||\mathrm{cum}(\epsilon_s)|
\end{aligned}$$

Both A and B reduce to a single summation because of the serial independence of the $\{\epsilon_k\}$, so $A = O(n^{-2d})$ and $B = O(n^{-2d})$. For C, by Surgailis and Viano (2002), Corollary 5.3,

$$|\mathrm{cum}(e^{h_k}, e^{h_s})| = e^{\sigma_h^2} |e^{r_{|k-s|}} - 1|$$

where $r_{|k-s|} = \mathrm{cov}(h_k, h_s)$ and $\sigma_h^2 = \mathrm{var}(h_k)$.

By the assumption on $\{b_j\}$ in the Theorem 2, it follows that $r_s \sim K s^{2d-1}$, as $s \to \infty$, where $K > 0$, so that

$$\begin{aligned}
\sum_{k=1}^{n} \sum_{s=1}^{n} |e^{r_{|k-s|}} - 1| &\leq 2 \sum_{k=1}^{n} \sum_{s>k}^{n} |e^{r_{|k-s|}} - 1| + n|e^{r_0} - 1| \\
&\leq Kn \sum_{j=1}^{n} j^{2d-1} + n|e^{r_0} - 1| = O(n^{2d+1}).
\end{aligned}$$

Thus term C is $O(1)$. Hence, $|\mathrm{cum}(y_n, y_n)|$ is $O(1)$.

Next, for the third order cumulant ($m = 3$), we have

$$|\mathrm{cum}(y_n, y_n, y_n)| = \frac{1}{n^{3d+\frac{3}{2}}} |\sum_{k=1}^{n} \sum_{s=1}^{n} \sum_{u=1}^{n} \mathrm{cum}(\tau_k, \tau_s, \tau_u)| \leq \frac{1}{n^{3d+\frac{3}{2}}} \sum_{k=1}^{n} \sum_{s=1}^{n} \sum_{u=1}^{n} |\mathrm{cum}(e^{h_k}\epsilon_k, e^{h_s}\epsilon_s, e^{h_u}\epsilon_u)| \quad .$$

We will use the following two-way table:

$$\begin{array}{cc}
e^{h_k} & \epsilon_k \\
e^{h_s} & \epsilon_s \\
e^{h_u} & \epsilon_u
\end{array}$$

For convenience, we group the indecomposable partitions according to how many sets ($L = 1, 2, 3$) the symbols $e^{h_k}, e^{h_s}, e^{h_u}$ are partitioned into.

We have three groups of indecomposable partitions, excluding those with at least one of $e^{h_k}, e^{h_s}, e^{h_u}$ and at least one of $\epsilon_k, \epsilon_s, \epsilon_u$ in the same set:

i) Group 1

$$\{e^{h_k}, e^{h_s}, e^{h_u}\}, \{\epsilon_k, \epsilon_s, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_s}, e^{h_u}\}, \{\epsilon_k, \epsilon_s\}, \{\epsilon_u\}$$

$$\{e^{h_k}, e^{h_s}, e^{h_u}\}, \{\epsilon_k, \epsilon_u\}, \{\epsilon_s\}$$

$$\{e^{h_k}, e^{h_s}, e^{h_u}\}, \{\epsilon_k\}, \{\epsilon_s, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_s}, e^{h_u}\}, \{\epsilon_k\}, \{\epsilon_s\}, \{\epsilon_u\}$$

ii) Group 2

$$\{e^{h_k}, e^{h_s}\}, \{e^{h_u}\}, \{\epsilon_k, \epsilon_s, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_s}\}, \{e^{h_u}\}, \{\epsilon_k\}, \{\epsilon_s, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_s}\}, \{e^{h_u}\}, \{\epsilon_s\}, \{\epsilon_k, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_u}\}, \{e^{h_s}\}, \{\epsilon_k, \epsilon_s, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_u}\}, \{e^{h_s}\}, \{\epsilon_k\}, \{\epsilon_s, \epsilon_u\}$$

$$\{e^{h_k}, e^{h_u}\}, \{e^{h_s}\}, \{\epsilon_s\}, \{\epsilon_k, \epsilon_u\}$$

$$\{e^{h_u}, e^{h_s}\}, \{e^{h_k}\}, \{\epsilon_k, \epsilon_s, \epsilon_u\}$$

$$\{e^{h_u}, e^{h_s}\}, \{e^{h_k}\}, \{\epsilon_k\}, \{\epsilon_s, \epsilon_u\}$$

$$\{e^{h_u}, e^{h_s}\}, \{e^{h_k}\}, \{\epsilon_s\}, \{\epsilon_k, \epsilon_u\}$$

iii) Group 3

$$\{e^{h_k}\}, \{e^{h_s}\}, \{e^{h_u}\}, \{\epsilon_k, \epsilon_s, \epsilon_u\}.$$

We next study the order of the dominant contribution to $|\mathrm{cum}(y_n, y_n, y_n)|$ corresponding to each group.

In Group 1, the dominant term arises from the last partition since it yields a triple summation,

$$\frac{1}{n^{3d+\frac{3}{2}}}\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n}|\mathrm{cum}(e^{h_k}, e^{h_s}, e^{h_u})||\mathrm{cum}(\epsilon_k)||\mathrm{cum}(\epsilon_s)||\mathrm{cum}(\epsilon_u)| = \frac{\mu_\epsilon^3}{n^{3d+\frac{3}{2}}}\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n}|\mathrm{cum}(e^{h_k}, e^{h_s}, e^{h_u})|$$

where $\mu_\epsilon = E^0(\epsilon_1)$.

By Surgailis and Viano (2002), Corollary 5.3,

$$\begin{aligned}
& \sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n}|\mathrm{cum}(e^{h_k}, e^{h_s}, e^{h_u})| \\
\leq\ & \sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} e^{\frac{3}{2}\sigma_h^2}|e^{r_{|k-s|}}-1||e^{r_{|k-u|}}-1||e^{r_{|s-u|}}-1| \\
+\ & \sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} e^{\frac{3}{2}\sigma_h^2}|e^{r_{|k-s|}}-1||e^{r_{|k-u|}}-1| + \sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} e^{\frac{3}{2}\sigma_h^2}|e^{r_{|k-s|}}-1||e^{r_{|s-u|}}-1| \\
+\ & \sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} e^{\frac{3}{2}\sigma_h^2}|e^{r_{|k-u|}}-1||e^{r_{|s-u|}}-1|
\end{aligned}$$

The last three summations are actually the same due to symmetry: we can simply relabel the indices in the last summation by $s \leftrightarrow u$. As for the first summation, since $|r_{|k-u|}| = |\mathrm{cov}(h_k, h_u)| \leq \sigma_h^2 = \mathrm{var}(h_k)$, we have $|e^{r_{|k-u|}} - 1| \leq (e^{\sigma_h^2} + 1) < \infty$. So

$$\begin{aligned}
\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n}|\mathrm{cum}(e^{h_k}, e^{h_s}, e^{h_u})| &\leq K\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} e^{\frac{3}{2}\sigma_h^2}|e^{r_{|k-s|}}-1||e^{r_{|s-u|}}-1| \\
&+ 3\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n} e^{\frac{3}{2}\sigma_h^2}|e^{r_{|k-s|}}-1||e^{r_{|s-u|}}-1| \\
&\leq K\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n}|e^{r_{|k-s|}}-1||e^{r_{|s-u|}}-1| \qquad \text{(for some } K > 0\text{)} \\
&= O(n^{4d+1})
\end{aligned}$$

The last step follows from Lemma 3. So $\frac{\mu_\epsilon^3}{n^{3d+\frac{3}{2}}}\sum_{k=1}^{n}\sum_{s=1}^{n}\sum_{u=1}^{n}|\mathrm{cum}(e^{h_k}, e^{h_s}, e^{h_u})|$ converges to zero because $(4d+1) < (3d+\frac{3}{2})$.

Similarly, the dominant contribution from Group 2 is of order

$$\frac{1}{n^{3d+\frac{3}{2}}}\sum_i\sum_j |\mathrm{cum}(e^{h_i}, e^{h_j})||\mathrm{cum}(e^{h_j})|$$

Note that in Group 2, all three of $e^{h_k}, e^{h_s}, e^{h_u}$ are partitioned into two sets. Therefore, partitions with all three of $\epsilon_k, \epsilon_s, \epsilon_u$ in different sets are not indecomposable, so the dominant contribution is a double sum,

$$\frac{1}{n^{3d+\frac{3}{2}}}\sum_i\sum_j |\mathrm{cum}(e^{h_i}, e^{h_j})||\mathrm{cum}(e^{h_j})| = \frac{\mu_{e^h}}{n^{3d+\frac{3}{2}}}\sum_i\sum_j |\mathrm{cum}(e^{h_i}, e^{h_j})| \le Kn^{(2d+1)-(3d+\frac{3}{2})} = O(n^{-d-\frac{1}{2}})$$

where $\mu_{e^h} = E^0(e^{h_1})$.

So the dominant term in Group 2 also converges to zero.

For Group 3, all three of $e^{h_k}, e^{h_s}, e^{h_u}$ are partitioned into three different sets, so that the part of the partition involving $\epsilon_k, \epsilon_s, \epsilon_u$ must be $\{\epsilon_k, \epsilon_s, \epsilon_u\}$ in order to be indecomposable. The resulting summation now is only a single one of order $O(n^1)$. The dominant contribution again converges to zero.

Notice that the order of the dominant contribution from group 3 ($O(n^{-3d-\frac{1}{2}})$) is of smaller order than that from group 2 ($O(n^{-d-\frac{1}{2}})$), which is of smaller order of that from group 1 ($O(n^{d-\frac{1}{2}})$). This will be shown to hold in general for any m-th order joint cumulant.

Next, we prove that the m-th order joint cumulant, which satisfies

$$|\mathrm{cum}(\underbrace{y_n, \ldots, y_n}_{m \text{ terms}})| \le \frac{1}{n^{m(d+\frac{1}{2})}}\sum_{k_1=1}^{n}\cdots\sum_{k_m=1}^{n} |\mathrm{cum}(e^{h_{k_1}}\epsilon_{k_1}, \ldots, e^{h_{k_m}}\epsilon_{k_m})| \tag{6}$$

converges to zero for all $m > 2$.

The indecomposable partitions of $(e^{h_{k_1}}\epsilon_{k_1}, \ldots, e^{h_{k_m}}\epsilon_{k_m})$ are organized in a similar manner as before into m groups, where in Group L the symbols $e^{h_{k_1}}, \ldots, e^{h_{k_m}}$ are divided into L sets ($L = 1, \ldots, m$).

a) First, consider Group 1. The dominant contribution to the righthand side of (6) corresponding to Group 1 must be the one from the partition in which all of the symbols $e^{h_{k_1}}, \ldots, e^{h_{k_m}}$ are in one set and each of the symbols $\epsilon_{k_1}, \ldots, \epsilon_{k_m}$ is in a set by itself. The resulting summation is an m-fold

summation. By Corollary 5.3 of Surgailis and Viano (2002), the absolute value of the m-th joint cumulant, $|\text{cum}(e^{h_{k_1}}, \ldots, e^{h_{k_m}})|$, is bounded by a summation taken over all connected graphs with m vertices. Each entry of the summation is a product of terms of the form $|e^{r_{|k_i-k_j|}} - 1|$ along the edges that connect vertices k_i and k_j of a connected m-vertex graph.

For a graph with m vertices, we need at least $(m-1)$ edges to connect them. It is known (see Andrasfai, 1977, Chapter 2) that any connected m-vertex graph with $(m-1)$ edges may be represented as a *tree*. Let $W_{\{k_i,\ldots,k_j\}} < \infty$ be the total number of trees with vertices labeled by $k_i, k_{i+1}, \ldots, k_j$.

If a connected m-vertex graph used in applying Corollary 5.3 of Surgailis and Viano (2002) has more than $(m-1)$ edges, it is not a tree, and there will be more than $(m-1)$ terms of the form $|e^{r_{|k_i-k_j|}} - 1|$ being multiplied together in the m-fold summation in (6). But, for all k_i, k_j, $|r_{|k_i-k_j|}| = |\text{cov}(h_{k_i}, h_{k_j})| \leq \sigma_h^2 = \text{var}(h_{k_i})$, so $|e^{r_{|k_i-k_j|}} - 1| \leq (e^{\sigma_h^2} + 1) < \infty$, and for any connected m-vertex graph with more than $(m-1)$ edges, there exists an m-vertex subgraph that has a tree representation. So we can retain a product of $(m-1)$ terms of the form $|e^{r_{|k_i-k_j|}} - 1|$ in the m-fold summation in (6) and move remaining terms out of the summation, bounding each by $(e^{\sigma_h^2} + 1)$. The resulting product of $(m-1)$ terms of the form $|e^{r_{|k_i-k_j|}} - 1|$ is itself a product over the edges of an m-vertex tree.

In all, $|\text{cum}(e^{h_{k_1}}, \ldots, e^{h_{k_m}})|$ is bounded by a constant times a summation over the set $G_{\{k_1,\ldots,k_m\}}$ of all $W_{\{k_1,\ldots,k_m\}}$ trees. Each entry of the summation is a product of terms of the form $|e^{r_{|k_i-k_j|}} - 1|$ being multiplied over the $(m-1)$ edges of the tree. Thus, we have

$$\begin{aligned}
\sum_{k_1=1}^{n} \cdots \sum_{k_m=1}^{n} |\text{cum}(e^{h_{k_1}}, \ldots, e^{h_{k_m}})| &\leq K \sum_{k_1=1}^{n} \cdots \sum_{k_m=1}^{n} \Big\{ \sum_{G_{\{k_1,\ldots,k_m\}}} \prod_{(k_i,k_j)\in\Omega(G_{\{k_1,\ldots,k_m\}})} |e^{r_{|k_i-k_j|}} - 1| \Big\}, \quad (K > 0) \\
&= K \sum_{G_{\{k_1,\ldots,k_m\}}} \sum_{k_1=1}^{n} \cdots \sum_{k_m=1}^{n} \Big\{ \underbrace{\prod_{(k_i,k_j)\in\Omega(G_{\{k_1,\ldots,k_m\}})} |e^{r_{|k_i-k_j|}} - 1|}_{(m-1)\text{ terms}} \Big\}
\end{aligned}$$

where $\Omega(G_{\{k_1,\ldots,k_m\}})$ is the set of edges of the graph indexed by $G_{\{k_1,\ldots,k_m\}}$.

By Lemma 3, each entry of the summation over $G_{\{k_1,\ldots,k_m\}}$ is of order $O(n^{2dm-2d+1})$. Also this

summation is taken over a finite number of graphs ($W_{\{k_1,\ldots,k_m\}} < \infty$), therefore

$$\sum_{k_1=1}^{n} \cdots \sum_{k_m=1}^{n} |\mathrm{cum}(e^{h_{k_1}}, \ldots, e^{h_{k_m}})| = O(n^{2dm-2d+1}).$$

Because the normalization term in (6) is of order $O(n^{m(d+\frac{1}{2})})$, the dominant contribution to $\mathrm{cum}(\underbrace{y_n, \ldots, y_n}_{m \text{ terms}})$ from Group 1 converges to zero, for any $m > 2$.

b) For Group 2, the symbols $e^{h_{k_1}}, \ldots, e^{h_{k_m}}$ are partitioned into two sets. Thus, the partitions with each of the m symbols $\epsilon_{k_1}, \ldots, \epsilon_{k_m}$ in a set by itself are not indecomposable. Relabel the two sets as $\{e^{h_{g_1}}, \ldots, e^{h_{g_q}}\}$, $\{e^{h_{g_{q+1}}}, \ldots, e^{h_{g_m}}\}$. Since the partition must be indecomposable, there must be one $I \in (1, \ldots, q)$ and one $J \in (q+1, \ldots, m)$, such that $g_I = g_J$. The dominant contribution to (6) from Group 2 is therefore

$$\frac{1}{n^{m(d+\frac{1}{2})}} \sum_{g_1=1}^{n} \cdots \sum_{g_m=1}^{n} |\mathrm{cum}(e^{h_{g_1}}, \ldots, e^{h_{g_q}})||\mathrm{cum}(e^{h_{g_{q+1}}}, \ldots, e^{h_{g_m}})||\mathrm{cum}(\epsilon_{g_I}, \epsilon_{g_J})| \tag{7}$$

Similarly as above, after applying Corollary 5.3 of Surgailis and Viano (2002) and after bounding certain terms, we obtain

$$\begin{aligned}
& \sum_{g_1=1}^{n} \cdots \sum_{g_m=1}^{n} |\mathrm{cum}(e^{h_{g_1}}, \ldots, e^{h_{g_q}})||\mathrm{cum}(e^{h_{g_{q+1}}}, \ldots, e^{h_{g_m}})||\mathrm{cum}(\epsilon_{g_I}, \epsilon_{g_J})| \\
\leq \;& K \sum_{g_1=1}^{n} \cdots \sum_{g_m=1}^{n} \Big\{ \sum_{G_{\{g_1,\ldots,g_q\}}} \underbrace{\prod_{(g_i,g_j)\in\Omega(G_{\{g_1,\ldots,g_q\}})} |e^{r_{|g_i-g_j|}} - 1|}_{(q-1) \text{ terms}} \Big\} \\
& \cdot \Big\{ \sum_{G_{\{g_{q+1},\ldots,g_m\}}} \underbrace{\prod_{(g_i,g_j)\in\Omega(G_{\{g_{q+1},\ldots,g_m\}})} |e^{r_{|g_i-g_j|}} - 1|}_{(m-q-1) \text{ terms}} \Big\} \Big\{ |\mathrm{cum}(\epsilon_{g_I}, \epsilon_{g_J})| \Big\} \\
= \;& K \sum_{G_{\{g_1,\ldots,g_q\}}} \sum_{G_{\{g_{q+1},\ldots,g_m\}}} \sum_{g_1=1}^{n} \cdots \sum_{g_m=1}^{n} \mathbf{1}_{\{g_I=g_J\}} \\
& \cdot \underbrace{\Big\{ \prod_{(g_i,g_j)\in\Omega(G_{\{g_1,\ldots,g_q\}})} |e^{r_{|g_i-g_j|}} - 1| \prod_{(g_i,g_j)\in\Omega(G_{\{g_{q+1},\ldots,g_m\}})} |e^{r_{|g_i-g_j|}} - 1| \Big\}}_{(m-2) \text{ terms, denote as } \Gamma(g_1,\ldots,g_m: G_{\{g_1,\ldots,g_q\}}, G_{\{g_{q+1},\ldots,g_m\}})}.
\end{aligned}$$

As mentioned before, any graph G_a in $G_{\{g_1,\ldots,g_q\}}$ and any graph G_b in $G_{\{g_{q+1},\ldots,g_m\}}$, can be represented by trees with q and $(m-q)$ vertices, respectively. Since for any two trees, the resulting structure obtained by merging one vertex from each tree is again a tree, under the constraint $g_I = g_J$, there exists a graph G_c in $G_{\{g_1,\ldots,g_{I-1},g_{I+1},\ldots,g_m\}}$, such that G_c is obtained by merging G_a and G_b together at the vertex $g_I = g_J$.

Therefore, the numerical value of the term Γ evaluated for graphs G_a and G_b and indices $\{g_1,\ldots,g_m\}$ with the constraint $g_I = g_J$ (which follows from the independence of the $\{\epsilon_{g_i}\}$) is equal to the value of the term Φ (defined below) evaluated using the graph G_c in $G_{\{g_1,\ldots,g_{I-1},g_{I+1},\ldots,g_m\}}$ and indices $\{g_1,\ldots,g_{I-1},g_{I+1},\ldots,g_m\}$ without any constraint on the values of these indices. After re-parameterizing $\{g_1,\ldots,g_{I-1},g_{I+1},\ldots,g_m\}$ by $\{l_1,\ldots,l_{m-1}\}$, we obtain

$$\begin{aligned}
& \sum_{g_1=1}^{n}\cdots\sum_{g_m=1}^{n} |\mathrm{cum}(e^{h_{g_1}},\ldots,e^{h_{g_q}})||\mathrm{cum}(e^{h_{g_{q+1}}},\ldots,e^{h_{g_m}})||\mathrm{cum}(\epsilon_{g_I},\epsilon_{g_J})| \\
\leq\ & K \sum_{G_{\{l_1,\ldots,l_{m-1}\}}} \sum_{l_1=1}^{n}\cdots\sum_{l_{m-1}=1}^{n} \underbrace{\prod_{(l_i,l_j)\in\Omega(G_{\{l_1,\ldots,l_{m-1}\}})} |e^{r|l_i-l_j|}-1|}_{(m-2)\text{ terms, denote as } \Phi(l_1,\ldots,l_{m-1}:G_{\{l_1,\ldots,l_{m-1}\}})} \\
=\ & O(n^{2d(m-2)+1})
\end{aligned}$$

where the final equality follows from Lemma 3.

The above $(m-1)$-fold summation for Group 2 is of smaller order than the m-fold summation from Group 1, which was $O(n^{2d(m-1)+1})$. Hence, the dominant contribution from Group 2 also converges to zero.

c) In general, for Group $L \in \{1,\ldots,m\}$, the symbols $e^{h_{k_1}},\ldots,e^{h_{k_m}}$ are partitioned into L sets. Relabel the L sets as $\{e^{h_{g_1}},\ldots,e^{h_{g_{q_1}}}\}$, $\{e^{h_{g_{q_1+1}}},\ldots,e^{h_{g_{q_2}}}\}$, $\ldots$, $\{e^{h_{g_{q_{L-1}+1}}},\ldots,e^{h_{g_m}}\}$. Since the partition must be indecomposable, there must be L indices $\{I,J,\ldots,Z\}$, where $I \in (1,\ldots,q_1)$, $J \in (q_1+1,\ldots,q_2),\ldots,Z \in (q_{L-1}+1,\ldots,m)$, such that $\underbrace{g_I = g_J = \ldots = g_Z}_{L\text{ terms}}$. The dominant contribution to (6) from Group L is then,

$$\frac{1}{n^{m(d+\frac{1}{2})}} \sum_{g_1=1}^{n}\cdots\sum_{g_m=1}^{n} \underbrace{|\mathrm{cum}(e^{h_{g_1}},\ldots,e^{h_{g_{q_1}}})|\ldots|\mathrm{cum}(e^{h_{g_{q_{L-1}+1}}},\ldots,e^{h_{g_m}})|}_{L-\text{terms}} |\mathrm{cum}(\underbrace{\epsilon_{g_I},\epsilon_{g_J},\ldots,\epsilon_{g_Z}}_{L\text{ terms}})|. \quad (8)$$

Similarly as before, we obtain

$$
\begin{aligned}
& \sum_{g_1=1}^{n} \cdots \sum_{g_m=1}^{n} \underbrace{|\mathrm{cum}(e^{h_{g_1}}, \ldots, e^{h_{g_{q_1}}})| \ldots |\mathrm{cum}(e^{h_{g_{q_{L-1}+1}}}, \ldots, e^{h_{g_m}})|}_{L-\text{terms}} \, |\mathrm{cum}(\underbrace{\epsilon_{g_I}, \epsilon_{g_J}, \ldots, \epsilon_{g_Z}}_{L \text{ terms}})| \\
\leq \; & K \underbrace{\sum_{G_{\{g_1,\ldots,g_{q_1}\}}} \cdots \sum_{G_{\{g_{q_{L-1}+1},\ldots,g_m\}}}}_{L-\text{fold}} \sum_{g_1=1}^{n} \cdots \sum_{g_m=1}^{n} \mathbf{1}\{\underbrace{g_I = g_J = \ldots = g_Z}_{L \text{ terms}}\} \\
& \cdot \underbrace{\Big\{ \prod_{(g_i,g_j)\in\Omega(G_{\{g_1,\ldots,g_{q_1}\}})} |e^{r_{|g_i-g_j|}} - 1| \quad \ldots \prod_{(g_i,g_j)\in\Omega(G_{\{g_{q_{L-1}+1},\ldots,g_m\}})} |e^{r_{|g_i-g_j|}} - 1| \Big\}}_{(m-L) \text{ terms}} \\
\leq \; & K \sum_{G_{\{l_1,\ldots,l_{m-L+1}\}}} \sum_{l_1=1}^{n} \cdots \sum_{l_{m-L+1}=1}^{n} \underbrace{\prod_{(l_i,l_j)\in\Omega(G_{\{l_1,\ldots,l_{m-L+1}\}})} |e^{r_{|l_i-l_j|}} - 1|}_{(m-L) \text{ terms}} \\
= \; & O(n^{2d(m-L)+1}),
\end{aligned}
$$

by Lemma 3.

The constraint $\underbrace{g_I = g_J = \ldots = g_Z}_{L \text{ terms}}$ allows the re-parameterization from $\{g_1, \ldots, g_m\}$ to $\{l_1, \ldots, l_{m-L+1}\}$ and reduces the m-fold summation in (8) to an $(m-L+1)$-fold summation in the last inequality. It was shown for Group 2 that the graph obtained by merging one vertex from each of any pair of trees is again a tree. By induction, we obtain a tree by merging one vertex from each of $L > 2$ trees, which allows us to apply Lemma 3 with $M = m - L + 1$ in the last step.

So, the dominant contribution from Group L is $O(n^{2d(m-L)+1-m(d+\frac{1}{2})})$, $(L = 1, \ldots, m)$. Since $d > 0$, the dominant contribution from *all* groups occurs for $L = 1$. Finally, the dominant contribution from Group 1 is $O(n^{2d(m-1)+1-m(d+\frac{1}{2})})$, which tends to zero for $m > 2$ since $d < \frac{1}{2}$. □

Lemma 2 *For durations $\{\tau_k\}$ satisfying the assumptions of Theorem 1,*

$$\limsup_{t} E[Z^4(t)] < \infty$$

where $Z(t)$ is defined by Equation (2).

Proof: By Chung (1974, Theorem 3.2.1, page 42), $E[Z(t)^4] \leq 1 + \sum_{s=1}^{\infty} P[Z^4(t) \geq s]$. Thus, it suffices to show that

$$\limsup_t \sum_{s=1}^{\infty} P[Z^4(t) \geq s] < \infty. \tag{9}$$

Note that for any real k,

$$N(t) \geq k \Longleftrightarrow \sum_{i=1}^{\lfloor k \rfloor} u_i \leq t. \tag{10}$$

We have

$$P[Z^4(t) \geq s] = P[Z(t) \leq -s^{1/4}] + P[Z(t) \geq s^{1/4}]. \tag{11}$$

Consider the second term $P[Z(t) \geq s^{1/4}]$. Using (10), we obtain

$$\begin{aligned} P[Z(t) \geq s^{1/4}] &= P[N(t) \geq \frac{t}{\mu} + s^{1/4} t^{1/2+d}] \\ &= P(\sum_{i=1}^{\lfloor g(t,s) \rfloor} u_i \leq t) \end{aligned}$$

where $g(t,s) = \frac{t}{\mu} + s^{1/4} t^{1/2+d}$.

So,

$$\begin{aligned} P[Z(t) \geq s^{1/4}] &= P(\sum_{i=1}^{\lfloor g(t,s) \rfloor} u_i \leq t) \\ &= P\Big(\frac{\sum_{i=1}^{\lfloor g(t,s) \rfloor}(u_i - \mu)}{\lfloor g(t,s) \rfloor^{1/2+d}} \leq \frac{t - \lfloor g(t,s) \rfloor \mu}{\lfloor g(t,s) \rfloor^{1/2+d}}\Big). \end{aligned}$$

Denote

$$U = \frac{\sum_{i=1}^{\lfloor g(t,s) \rfloor}(u_i - \mu)}{\lfloor g(t,s) \rfloor^{1/2+d}} \quad .$$

Since $-\lfloor x\rfloor < -x+1$ for $x>0$, we obtain for any positive p,

$$\begin{aligned} P[Z(t) \geq s^{1/4}] &= P\Big(U \leq \frac{t-\lfloor g(t,s)\rfloor\mu}{\lfloor g(t,s)\rfloor^{1/2+d}}\Big) \\ &\leq P\Big(U \leq \frac{-\mu s^{1/4}t^{1/2+d}+\mu}{\lfloor g(t,s)\rfloor^{1/2+d}}\Big) \\ &\leq P\Big(|U| \geq \frac{\mu s^{1/4}t^{1/2+d}-\mu}{\lfloor g(t,s)\rfloor^{1/2+d}}\Big) \\ &\leq KE(|U|^p)\frac{(\frac{t}{\mu}+s^{1/4}t^{1/2+d})^{(1/2+d)p}}{\mu^p[s^{1/4}t^{1/2+d}-1]^p} \end{aligned}$$

For $t \geq 4$, since $s^{1/4}t^{1/2+d}-1 \geq \frac{1}{2}s^{1/4}t^{1/2+d}$ and $\frac{1}{2}+d<1$, we obtain

$$P[Z(t) \geq s^{1/4}] \leq KE(|U|^p)\frac{1}{s^{\frac{p}{4}(\frac{1}{2}-d)}} \tag{12}$$

Now, consider

$$P[Z(t) \leq -s^{1/4}] = P[N(t) \leq \frac{t}{\mu} - s^{1/4}t^{1/2+d}] \quad .$$

Let $a(t) = \frac{t^{2-4d}}{\mu^4}$ and $v(t,s) = \frac{t}{\mu} - s^{1/4}t^{1/2+d}$. Using (10), we have

$$\begin{aligned} P[Z(t) \leq -s^{1/4}] &= P(\sum_{i=1}^{\lfloor v(t,s)\rfloor} u_i > t), \qquad s<a(t) \\ &= P(u_1>t), \qquad s=a(t) \\ &= 0, \qquad s>a(t) \end{aligned}$$

For $s<a(t)$, we have $v(t,s)>0$. Let

$$W = \frac{\sum_{i=1}^{\lfloor v(t,s)\rfloor}(u_i-\mu)}{\lfloor v(t,s)\rfloor^{1/2+d}} \quad .$$

Then for any positive p,

$$\begin{aligned} P[Z(t) \leq -s^{1/4}] = P[N(t) \leq v(t,s)] &= P\Big(\frac{\sum_{i=1}^{\lfloor v(t,s)\rfloor}(u_i-\mu)}{\lfloor v(t,s)\rfloor^{1/2+d}} > \frac{t-\mu\lfloor v(t,s)\rfloor}{\lfloor v(t,s)\rfloor^{1/2+d}}\Big) \\ &\leq P\Big(W > \frac{\mu s^{1/4}t^{1/2+d}}{(\frac{t}{\mu}-s^{1/4}t^{1/2+d})^{1/2+d}}\Big) \\ &\leq KE(|W|^p)\frac{(\frac{t}{\mu}-s^{1/4}t^{1/2+d})^{(1/2+d)p}}{s^{p/4}t^{(1/2+d)p}} \\ &\leq KE(|W|^p)\frac{1}{s^{p/4}} \end{aligned}$$

i.e.

$$P[Z(t) \leq -s^{1/4}] \leq KE(|W|^p)\frac{1}{s^{p/4}} \quad . \tag{13}$$

For $s = a(t)$, $P[Z(t) \leq -s^{1/4}] = P[u_1 > t] \leq \frac{E(u_1)}{t}$.

For $s > a(t)$, $P[Z(t) \leq -s^{1/4}] = 0$.

Select any positive p such that $\frac{p}{4}(\frac{1}{2} - d) > 1$ and thus $\frac{p}{4} > 1$ since $0 < \frac{1}{2} - d < 1$. If it can be shown that $\sup_{t>1,s>1} E(|U|^p) < \infty$ and $\sup_{t>1,s>1} E(|W|^p) < \infty$, then by (12) and (13), it follows that $P[Z^4(t) \geq s]$ is summable, uniformly in t. Thus, (9) follows and the proof is complete.

We next show that indeed $\sup_{t>1,s>1} E(|U|^p) < \infty$ and $\sup_{t>1,s>1} E(|W|^p) < \infty$ for all positive p when $d \in (0, \frac{1}{2})$ and for $p = 8 + \delta, \delta > 0$ when $d = 0$. Define

$$B_1 = \frac{u_1 - \mu}{\lfloor g(t,s) \rfloor^{1/2+d}} \quad , \quad B_2 = \frac{\sum_{i=2}^{\lfloor g(t,s) \rfloor} (\tau_i - \mu)}{\lfloor g(t,s) \rfloor^{1/2+d}},$$

so that $U = B_1 + B_2$. By Minkowski's Inequality,

$$E[|U|^p] \leq \left[(E|B_1|^p)^{1/p} + (E|B_2|^p)^{1/p} \right]^p \quad .$$

Since $u_1 \leq \tau_1$, using $h(x) = (x + \mu)^p$ in (1), and since by assumption iv), τ_1 has all finite moments up to order p under P^0, we have

$$\sup_{t>1,s>1} E|B_1|^p < \infty \quad .$$

From Baccelli and Brémaud (2003, Equation 1.2.25, page 20) that for any measurable function h,

$$E[h(\tau_2, \ldots, \tau_n)] = \lambda E^0[\tau_1 h(\tau_2, \ldots, \tau_n)] \quad .$$

This, together with the Cauchy-Schwarz inequality, yields

$$E|B_2|^p = \lambda E^0(\tau_1 |B_2|^p) \leq \lambda [E^0(\tau_1^2)]^{1/2} \, [E^0 |B_2|^{2p}]^{1/2} \quad ,$$

where $\lambda = 1/E^0(\tau_1)$. By assumption iv), $\sup_{t>1,s>1} E^0|B_2|^p < \infty$. for all positive p when $d \in (0, \frac{1}{2})$ and for $p = 8 + \delta, \delta > 0$ when $d = 0$. It follows that $\sup_{t>1,s>1} E[|U|^p] < \infty$. By a similar argument, $\sup_{t>1,s>1} E[|W|^p] < \infty$. $\square$

Lemma 3 *For any $M > 2$ and $0 < d < \frac{1}{2}$,*

$$\underbrace{\sum_{k_1=1}^{n} \cdots \sum_{k_M=1}^{n}}_{M-fold} \underbrace{\left\{ \prod_{(k_i,k_j)\in\Omega(G)} |e^{r_{|k_i-k_j|}} - 1| \right\}}_{(M-1)\ terms} = O(n^{2d(M-1)+1}) \tag{14}$$

where $\Omega(G)$ is the set of edges of G, G is any connected M-vertex graph with vertices $\{k_1, \ldots, k_M\}$ and $(M-1)$ edges; $r_{|k_i-k_j|} = cov(h_{k_i}, h_{k_j}), 1 \le i \le M, 1 \le j \le M$, $\{h_{k_i}\}$ is a long memory process with memory parameter d.

Proof: Since G is a connected graph with M vertices and $(M-1)$ edges, it can be represented as a tree (see Andrasfai 1977, Chapter 2). The tree representation is not unique. Fix a particular representation. Then there is one vertex with no parent, called the root. A vertex with both a parent and a child is called a node. A vertex with no child is called a leaf.

We proceed iteratively. First, select any leaf vertex. By definition of a leaf, the corresponding index only appears once in the product, so the sum on this index can be evaluated for this term only, holding the other terms fixed. Since $r_s \sim Cs^{2d-1}$ as $s \to \infty$, we have for any fixed integer i with $1 \le i \le n$, $\sum_{j=1}^{n} |e^{r_{|i-j|}} - 1| = O(n^{2d})$.

It follows that the sum on the first index is $O(n^{2d})$. Next, delete the leaf just used from the tree. The resulting graph is again a tree. Repeat the process of selecting a leaf, performing the corresponding sum and deleting the leaf until only the root remains. The M-fold sum in (14) is now bounded by a constant times the sum of n terms each of which is $O(n^{2d(M-1)})$. Thus, the sum in (14) is $O(n^{2d(M-1)+1})$. $\square$

Lemma 4 *Under the LMSD model described in Theorem 2 with memory parameter $d \in [0, \frac{1}{2})$, P^0 is $\{\tau_k\}$-mixing; The durations $\{\tau_k\}$ generated by the ACD(1,1) model described in Theorem 3 are exponential α-mixing.*

Proof: Under P^0, $\{h_k\}$ is a stationary Gaussian process with a log spectral density having an integral on $[-\pi, \pi]$ that is greater than $-\infty$, so that the innovation variance is positive. Since Gaussian processes

are time reversible, it follows that we can represent $h_k = \sum_{j=0}^{\infty} a_j w_{k+j}$ where $\sum a_j^2 < \infty$ and $\{w_k\}$ is an *iid* Gaussian sequence. Arguing as in the proof of Theorem 17.3.1 of Ibragimov and Linnik (1971), *pp.* 311–312, replacing $\{\ldots w_{k-1}, w_k\}$ by $\{w_k, w_{k+1}, \ldots\}$, it follows that P^0 is $\{h_k\}$-mixing. Since the $\{\epsilon_k\}$ are *iid* it follows that P^0 is also $\{\epsilon_k\}$-mixing. Since for any process $\{\xi_k\}$, P^0 is $\{\xi_k\}$-mixing if and only if the future tail σ-field of $\{\xi_k\}$ is trivial (see, e.g., Nieuwenhuis (1989), Equation (3.3)), it follows from Lemma 5 that P^0 is $\{\tau_k\}$-mixing, where $\tau_k = e^{h_k}\epsilon_k$.

For the ACD(1,1) model, by Proposition 17 of Carrasco and Chen (2002), $\{\tau_k\}$ is exponential β-mixing (or also called absolutely regular) if $\{\tau_0, \psi_0\}$ are initialized from the stationary distribution. Their result still holds for a doubly infinite sequence $\{\tau_k\}, k \in (-\infty, \infty)$. It is well known that β-mixing implies α-mixing (or strong mixing), (see Bradley (2005), Section 2.1). Therefore, $\{\tau_k\}$ is also exponential α-mixing, which further implies $\{\tau_k\}$-mixing of P^0 for the ACD(1,1) model, see Nieuwenhuis (1989), Equation (3.5). $\square$

Lemma 5 *Let $\{\xi_s\}$ and $\{\zeta_s\}$ be two* independent *processes whose future tail σ-fields are trivial. Then the future tail σ-field of the process $\{\xi_s, \zeta_s\}$ is trivial.*

Proof: Define $\mathcal{S}_t = \sigma(\xi_s, s \geq t)$, $\mathcal{T}_t = \sigma(\zeta_s, s \geq t)$ and $\mathcal{U}_t = \sigma(\xi_s, \zeta_s, s \geq t)$. As pointed out by Ibragimov and Linnik (1971, p. 303) (for regularity), to prove that $\mathcal{U}_\infty$ is trivial, it suffices to prove that for all $\mathcal{U}_0$-measurable zero mean random variables η such that $\mathbf{E}[\eta^2] \leq 1$, $\mathbf{E}[\eta \mid \mathcal{U}_t]$ converges to 0 in quadratic mean. By standard arguments, it suffices to prove this for a random variable η that can be expressed as $\eta = \eta_1\eta_2$ with η_1 $\mathcal{S}_0$-measurable and η_2 $\mathcal{T}_0$-measurable and, without loss of generality, both with zero mean. Then, by independence of $\{\xi_s\}$ and $\{\zeta_s\}$,

$$\mathbf{E}[\eta \mid \mathcal{U}_t] = \mathbf{E}[\eta_1 \mid \mathcal{S}_t] \times \mathbf{E}[\eta_2 \mid \mathcal{T}_t] \; .$$

Since $\mathcal{S}_\infty$ and $\mathcal{T}_\infty$ are trivial, both terms in the right hand side above tend to 0 in q.m. By independence, their product also tends to 0 in q.m. $\square$

References

[1] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001), The Distribution of Exchange Rate Volatility, *Journal of the American Statistical Association* **96**, 42–55.

[2] Andrasfai, B. (1977), *Introductory Graph Theory.* Bristol, UK: Adam Hilger.

[3] Baccelli, F. and Brémaud, P. (2003), *Elements of Queueing Theory.* New York: Springer.

[4] Bollerslev, T. and Mikkelsen, H.O. (1996), Modeling and Pricing Long Memory in Stock Market Volatility, *Journal of Econometrics* **73**, 151–184.

[5] Bradley, R. (2005), Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions, *Probability Surveys* **2**, 107–144.

[6] Brillinger, D.R. (1981), *Time Series. Data Analysis and Theory*, 2'nd Ed. Oakland, CA: Holden-Day.

[7] Carrasco, M. and Chen X. (2002), Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models, *Econometric Theory* **18**, 17–39.

[8] Chung, K.L. (1974), *A Course In Probability Theory*, 2'nd Ed. New York: Academic Press.

[9] Comte, F. and Renault, E. (1998), Long Memory in Continuous-Time Stochastic Volatility Models, *Mathematical Finance* **8**, 291–323.

[10] Daley, D.J., Rolski, T. and Vesilo, R. (2000), Long-Range Dependent Point Processes and their Palm-Khinchin Distributions, *Advances in Applied Probability* **32**, 1051–1063.

[11] Daley, D.J. and Vere-Jones, D. (2003), *An Introduction to the Theory of Point Processes*, 2'nd Ed. New York: Springer.

[12] Deo, R.S. and Hurvich, C.M. (2001), On the Log Periodogram Regression Estimator of the Memory Parameter in Long Memory Stochastic Volatility Models, *Econometric Theory* **17**, 686–710.

[13] Deo, R. Hsieh, M. and Hurvich, C.M. (2005), Tracing the Source of Long Memory in Volatility. Preprint, New York University Stern School of Business.

[14] Deo, R., Hurvich, C. and Lu, Y. (2005), Forecasting Realized Volatility Using a Long Memory Stochastic Volatility Model: Estimation, Prediction and Seasonal Adjustment. To appear in *Journal of Econometrics*.

[15] Doukhan, P. (1994), *Mixing*, Lecture Notes in Statistics **85**, New York: Springer-Verlag.

[16] Engle, R. and Russell J. (1998), Autoregressive Conditional Duration: a New Model for Irregularly Spaced Transaction Data, *Econometrica* **66**, 1127–1162.

[17] Guégan, D. and Ladoucette, S. (2001), Non-Mixing Properties of Long Memory Processes, *C.R. Acad. Sci. Paris Sér I Math.* **333**, 373–376.

[18] Hurvich, C.M., Moulines, E. and Soulier, P. (2005), Estimating Long Memory in Volatility, *Econometrica* **73**, 1283–1328.

[19] Ibragimov, I.A. and Linnik, Yu.V. (1971), *Independent and Stationary Sequences of Random Variables*, Ed. J.F.C. Kingman, Groningen: Wolters-Noordhoff Publishing.

[20] Iglehart, D.L. and Whitt, W. (1971), The Equivalence of Functional Central Limit Theorems for Counting Processes and Associated Partial Sums, *Annals of Mathematical Statistics* **42**, 1372–1378.

[21] Nelson, D. (1990), Stationarity and Persistence in the GARCH(1,1) Model, *Econometric Theory* **6**, 318–334.

[22] Nieuwenhuis, G. (1989), Equivalence of Functional Limit Theorems for Stationary Point Processes and Their Palm Distributions, *Probability Theory and Related Fields* **81**, 593–608.

[23] Robinson, P.M. (1991), Testing for Strong Serial Correlation and Dynamic Conditional Heteroskedasticity in Multiple Regression, *Journal of Econometrics* **47**, 67–84.

[24] Robinson, P.M. and Henry, M. (1999), Long and Short Memory Conditional Heteroskedasticity in Estimating the Memory Parameter of Levels, *Econometric Theory* **15**, 299–336.

[25] Surgailis, D. and Viano, M-C. (2002), Long Memory Properties and Covariance Structure of the EGARCH Model, *European Series in Applied and Industrial Mathematics: Probability and Statistics* **6**, 311–329.

[26] Tsay, R. (2002), *Analysis of Financial Time Series*, New York: John Wiley & Sons.

[27] Yokoyama, R. (1980), Moment Bounds for Stationary Mixing Sequences, *Z. Wahrsch. Verw. Gebiete* **52**, 45–57.